KARIN M. VERSPOOR

http://public.lanl.gov/verspoor
(505) 667-5086 (office) verspoor@lanl.gov

SUMMARY OF QUALIFICATIONS

Versatile computational linguistics researcher with particular expertise in methods for representing and manipulating semantic information, utilizing hybrid methodologies combining statistical and quantitative analysis with symbolic, logical, and qualitative techniques. Hands-on development experience building natural language processing systems in the private sector as well as for government customers.

PROFESSIONAL EXPERIENCE

LOS ALAMOS NATIONAL LABORATORY – *Los Alamos, New Mexico* **2003 - present**

Technical Staff Member, Knowledge and Information Systems Science team

Chosen to perform scientific research in computational methods for the extraction, representation, organization, synthesis, discovery and retrieval of knowledge in databases and information systems.

- Delivered information extraction systems for bioinformatics and homeland security applications, through integration of existing third-party tools and newly developed task-specific components.
- Submitted state of the art results to open evaluations in the areas of information extraction from biological text and protein function prediction using systems developed to take advantage of the mathematical properties of biomedical resources (databases and ontologies).
- Awarded a Los Alamos Exploratory Research grant to investigate unsupervised methods for induction of patterns for information extraction.
- Identified and validated new unsupervised methods for word sense disambiguation in collaboration with a postdoctoral research assistant under my supervision.
- Streamlined parallel text processing efforts across four national laboratories for several homeland security applications by gaining adoption of a common architecture and minimizing redundancies in the efforts.
- Invited to participate in standards definitions for unstructured information management through OASIS-Open, building on IBM's Unstructured Information Management Architecture (UIMA).

APPLIED SEMANTICS, INC. – *Los Angeles, California* **2001 - 2002**

Senior Computational Linguist

Brought on to lead core technology development of an enterprise suite offering semantic meta-tagging, summarization and categorization of documents.

- Achieved improvements in system quality via redesign of existing word sense disambiguation and "page sensing" algorithms to incorporate NLP techniques such as reference resolution and shallow parsing.
- Augmented lexicon and ontology specificity through implementation of statistical techniques for semantic clustering and automatic acquisition of concept-contexts and selectional preferences.
- Supported a successful patent filing (US Patent # 6,453,315) by rewriting the patent description to clarify the technology and to reference appropriate literature, and serving as attorney's technical point of contact.

WEBMIND, INC. – *New York, New York* **1998 - 2001**

Director, Natural Language Engineering

Hired to design and implement the natural language module in the Webmind Artificial Intelligence system, a module aimed at deep semantic understanding of texts and intelligent, conversational query answering.

- Directed a team of nine scientists and programmers, leading conceptual and implementation-level design.
- Delivered a natural language processing module capable of fine-grained semantic representation of text, through integration of techniques derived from information retrieval algorithms, statistical parsing, unification-based analysis and information extraction.
- Created an approach to automated language acquisition called "experiential language learning" in collaboration with the scientists developing the general intelligence dynamics of the Webmind architecture.

KARIN M. VERSPOOR

http://public.lanl.gov/verspoor
(505) 667-5086 (office) verspoor@lanl.gov

MICROSOFT RESEARCH INSTITUTE – *Sydney, Australia* **1997 - 1998**
Research Fellow
Selected to conduct research in multilingual natural language generation, and to serve as Project Coordinator of the DDD (Dynamic Document Delivery) project.

- Delivered a system to the Powerhouse Museum in Sydney, Australia that generated multilingual textual descriptions of museum objects from an abstract database at varying levels of detail.
- Liased with museum representatives to gain access to and understand their data, and to deploy the system to the museum website.
- Supervised linguistic experts in Spanish and Chinese to develop the multilingual resources of the system, redesigning a base English-only system to accommodate the multilinguality.

TECHNICAL EXPERTISE

Computational linguistics research experience:

- Information extraction, ontology-based methods for semantic analysis, lexical semantics, word sense disambiguation, coreference resolution and natural language generation.
- Language modeling, latent semantic analysis and terminology induction.
- Extensive background in theoretical linguistics and formal semantics.

Extensive software engineering and data management experience, including:

- Programming languages: Java, Perl, C++, Lisp, UNIX scripting tools, some Python.
- Database systems: Using ORACLE, DB2, SQL Server for data storage and analysis.

EDUCATION

UNIVERSITY OF EDINBURGH – *Edinburgh, Scotland, UK* **1997**
PhD, Cognitive Science and Natural Language Thesis: "Contextually-Dependent Lexical Semantics"

UNIVERSITY OF EDINBURGH – *Edinburgh, Scotland, UK* **1994**
MSc, Cognitive Science and Natural Language Thesis: "A Cognitively-Relevant Lexical Semantics"
Awarded with Distinction

RICE UNIVERSITY – *Houston, TX* **1993**
BA, Computer Science and Cognitive Sciences Thesis: "What are the Characteristics of Emotional Metaphors?"
Graduated *Summa Cum Laude*

PROFESSIONAL SERVICE

- Los Alamos National Laboratory, Women's Diversity Working Group, co-chair 2006-present
- Los Alamos National Laboratory, Diversity/Affirmative Action Board, Secretary 2005-2007
- Los Alamos National Laboratory, Exploratory Research Grants Computer Science review committee, 2006
- Chair, BioNLP'06. Linking Natural Language Processing and Biology, Workshop at NAACL 2006
- Chair, UCLA IPAM Workshop on Dynamic Searches and Knowledge Building, October 2007
- Programme Committees and reviewing for numerous conferences, workshops, and journals
- Teaching and tutoring in computer science and computational linguistics 1991-1998
- Adjunct Professor, Communication Studies Department, New Mexico State University, 2005-2006

KARIN M. VERSPOOR

http://public.lanl.gov/verspoor
(505) 667-5086 (office) verspoor@lanl.gov

SELECTED PUBLICATIONS

- Lin, S. and Verspoor, K. (2007). Learning Semantics-Enhanced Language Models Applied to Unsupervised Word Sense Disambiguation. Submitted.
- Verspoor, K., Sanfilippo, A., Elmore, M., MacKerrow, E. (2006). Deploying Natural Language Processing for Social Science Analysis. In proceedings "What to Do with a Million Books": Chicago Colloquium on Digital Humanities and Computer Science.
- Verspoor, K., Cohn, J., Mniszewski, S., and Joslyn, C. (2006). A Categorization Approach to Automated Ontological Function Annotation. *Protein Science*, v.15, pp.1544-1549.
- Joslyn, C., Gessler, D.D.G., Schmidt, S.E., Verspoor, K.M. (2006) Distributed Representations of Bio-Ontologies for Semantic Web Services. In proceedings Joint BioLink and BioOntologies SIG of ISMB 2006.
- Maguitman, A., Rechtsteiner, A., Verspoor, K., Strauss, C.E., Rocha, L. (2006) Large-Scale Testing Of Bibliome Informatics Using Pfam Protein Families. Pacific Symposium of Biocomputing 11:76-87
- Verspoor, Karin. (2005). Towards a semantic lexicon for biological language processing. *Comparative and Functional Genomics*, vol. 6, issue 1-2, p. 61-66. DOI: 10.1002/cfg.451.
- Verspoor, K., Cohn, J., Mniszewski, S., Joslyn, C. (2005). POSOLE: Automated Ontological Annotation for Function Prediction. In proceedings, Automated Function Prediction SIG at ISMB 2005.
- Joslyn, C., Cohn, J., Verspoor, K., Mniszewski, S. (2005) Automating Ontological Function Annotation: Towards a Common Methodological Framework. In proceedings, BioOntologies SIG at ISMB 2005.
- Verspoor, K., Cohn, J., Joslyn, C., Mniszewski, S., Rechtsteiner, A., Rocha, L.M., Simas, T. (2005). Protein Annotation as Term Categorization in the Gene Ontology using Word Proximity Networks. *BMC Bioinformatics* 2005, vol. 6 (suppl 1).
- Verspoor, C., C. Joslyn and G. Papcun (2003a). "Interactions Between the Gene Ontology and a Domain Corpus for a Biological Natural Language Processing Application". In proceedings Sixth Annual Bio-Ontologies Meeting.
- Verspoor, C., C. Joslyn and G. Papcun (2003b). "The Gene Ontology as a Source of Lexical Semantic Knowledge for a Biological Natural Language Processing Application". In proceedings SIGIR'03 Workshop on Text Analysis and Search for Bioinformatics.
- Robert Dale, Stephen J Green, Maria Milosavljevic, Cecile Paris, Cornelia Verspoor and Sandra Williams (1998). Dynamic Document Delivery: Generating Natural Language Texts on Demand. Presented at the 9th International Conference and Workshop on Database and Expert Systems Applications (DEXA'98), University of Vienna.
- Wan, S. and Verspoor, C. (1998), "Automatic English-Chinese name transliteration for development of multilingual resources." In proceedings of COLING-ACL'98, Montreal, Canada.
- Verspoor, C., Dale, R., Green, S., Milosavljevic, M., Paris, C., and Williams, S. (1998) "Intelligent Agents for Information Presentation: Dynamic Description of Knowledge Base Objects". In proceedings International Workshop on Intelligent Agents on the Internet and Web, Mexico City, Mexico, pp. 75-86.
- Dale, R., Green, S., Milosavljevic, M., Paris, C., Verspoor, C. and Williams, S. (1998). "Using Natural Language Generation Techniques to Produce Virtual Documents", In proceedings Third Australian Document Computing Symposium (ADCS'98).
- Dale, R., Green, S., Milosavljevic, M., Paris, C., Verspoor, C., Williams, S. (1998). "The Realities of Generating Natural Language from Databases". In proceedings 11th Australian Joint Conference on Artificial Intelligence.
- Verspoor, C.M. (1998) "The author explores the issue onto the agenda: Predictivity vs. Stipulativity in the Lexicon" In proceedings Pacific Asia Conference on Language, Information, and Computation, pp. 152-162.
- Verspoor, C.M. (1997) "Lexical Resources for Natural Language Processing Systems and the Problem of Polysemy". In poster proceedings Tenth Australian Joint Conference on Artificial Intelligence, pp. 150-155.
- Verspoor, C.M. (1997) "Conventionality-Governed Logical Metonymy". In proceedings 2nd International Workshop on Computational Semantics, H. Bunt, L. Kievit, R. Muskens and M. Verlinden, eds.,pp. 300-312.
- Narasimhan, B., Di Tomaso, V. and Verspoor, C.M. (1996) "Unaccusative or Unergative? Verbs of Manner of Motion". Quaderni del laboratorio di linguistica, 10, Scuola Normale Superiore of Pisa.
- Verspoor, C.M. (1996a) "A Perspective on PPs" Edinburgh Working Papers in Cognitive Science, Vol. 12: Studies in Head-Driven Phrase Structure Grammar, Claire Grover and Enric Vallduvi, eds., pp. 229-271.
- Verspoor, C.M. (1996b) "Lexical Limits on the Influence of Context" In proceedings Eighteenth Annual Conference of the Cognitive Science Society. Lawrence Erlbaum Associates Publishers, pp. 116-120.